

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2001-92481

(P2001-92481A)

(43)公開日 平成13年4月6日(2001.4.6)

(51)Int.Cl.<sup>7</sup>

G 1 0 L 13/06  
13/08

識別記号

F I

G 1 0 L 5/04  
3/00  
5/04

テ-マコ-ト\*(参考)

F 5 D 0 4 5  
H  
B

審査請求 未請求 請求項の数 3 O L (全 5 頁)

(21)出願番号

特願平11-269884

(22)出願日

平成11年9月24日(1999.9.24)

(71)出願人 000001889

三洋電機株式会社

大阪府守口市京阪本通2丁目5番5号

(72)発明者 平井 啓之

大阪府守口市京阪本通2丁目5番5号 三  
洋電機株式会社内

(72)発明者 橋本 誠

大阪府守口市京阪本通2丁目5番5号 三  
洋電機株式会社内

(74)代理人 100086391

弁理士 香山 秀幸

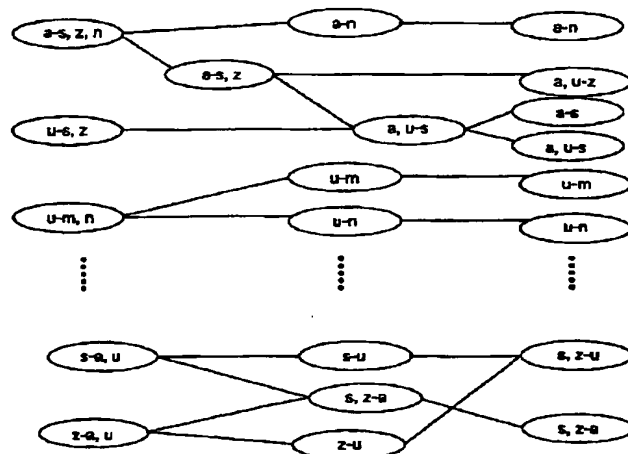
最終頁に続く

(54)【発明の名称】 規則音声合成方法

(57)【要約】

【課題】 この発明は、予め用意したCV、VC単位の素片を利用して音声合成を行なう音声合成方法であって、音声辞書のサイズの低減させることができるとともに接続面での歪みを減少させることができる規則音声合成方法を提供することを目的とする。

【解決手段】 この発明による規則音声合成方法は、H M n e tを用いてCV、VC単位の素片を、音素単位より細かい合成単位に分割し、全ての合成単位にまたがる音素内ネットワークを生成し、得られた音素内ネットワークに基づいて入力音素列に対応する素片を選択するようにしたことを特徴とする。



## 【特許請求の範囲】

【請求項1】 HMnetを用いてCV、VC単位の素片を、音素単位より細かい合成単位に分割し、全ての合成単位にまたがる音素内ネットワークを生成し、得られた音素内ネットワークに基づいて入力音素列に対応する素片を選択するようにした規則音声合成方法。

【請求項2】 音素内ネットワークは、予め用意した個々のCV、VC単位の素片の始端部分から得られる複数の第1状態と、全てのCV、VC単位の素片の中央部分から得られる1つの第2状態と、個々のCV、VC単位の素片の終端部分から得られる複数の第3状態とによって構成されるHMnetの初期状態から、逐次状態分割法によって状態を分割するステップ、および任意の状態数まで分割されたHMnetと、個々のCV、VC単位の素片とに基づいて、HMnetの各状態の代表素片を決定するステップ、によって生成される請求項1に記載の規則音声合成方法。

【請求項3】 HMnetの各状態の代表素片を決定するステップは、個々のCV、VC単位の素片と、任意の状態数まで分割されHMnetから、各学習用素片を各状態に分割し、各状態の素片の候補を作成するステップ、およびHMnetの全ての状態間の接続点に、HMnetの同じ音素からなる終端と始端の全ての組み合わせを加えたものを総接続点とし、総接続点での接続歪みの総和が最小となるような素片の組み合わせを各状態の素片候補の中から選択して、代表素片とするステップ、を備えている請求項2に記載の規則音声合成方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】この発明は、規則音声合成方法に関する。

## 【0002】

【従来の技術】規則音声合成では、一般に、音声データから抽出した接続単位（音声素片）を、合成したい文章にあわせて選択し、韻律を修正して接続することによって音声合成が行なわれる。

【0003】接続単位には、(C、V)、(CV、VC)、VCV等が用いられる。ここで、Cは子音を示し、Vは母音を示している。

【0004】接続単位数が最も少ないのは、音素(C、V)を接続単位とする場合であるが、音素と音素との境界では、前後の音素の種類によって音響的性質が異なるため、各音素毎に1つの素片しか持たないとすると、接続面での歪みが増加する。このため、一般的に、各音素毎に複数の素片を持つ必要がある。各接続単位毎に複数の素片を持たせるようにすると、音質は改善されるが、音声辞書のサイズを増大させることになる。

【0005】ところで、(CV、VC)を接続単位にし

た場合、日本語を合成しようとする、各接続単位毎に1つの素片しか持たせないとすると、500程度の素片が必要となる。さらに、音質の面から見ても、各接続単位毎に1つの素片しか持たないとすると、先行または後続音素が異なる接続単位を接続する場合が生じ、大きな接続歪みが発生する原因となる。

【0006】(CV、VC)を接続単位とした場合、異なる接続単位間でも同じような波形が重複して含まれている。このような部分を削除するためには、接続単位をより細かくし、同じような音響的性質の部分を1つの接続単位（合成単位）として共通化すればよい。また、先行および後続音素が異なることによる影響は、音素全てに均等に現れるのではない。したがって、接続単位が細かければ、影響の大きい部分にだけ効率よく接続単位を増やすことができる。

【0007】このような考え方は、音素(C、V)を接続単位とした場合でも成り立ち、既に伊藤ら（特開平9-222898参照）によって提案されている。つまり、音素毎にノードを介して複数のアークが連結され、各アークに対応する音響特徴パラメータが添付されている音響パラメータネットワークを記憶しておき、入力音素列に応じてアークの系列を探索し、音響特徴パラメータ列を形成し、音声の合成を行なうものである。

【0008】この方法では各音素毎にネットワークを形成している。一方、予め用意したCV、VC単位の素片からなる音声データベースに対しては、全ての合成単位にまたがるネットワークを構成する必要があるため、簡単にはこの方法を応用できない。

## 【0009】

【発明が解決しようとする課題】この発明は、予め用意したCV、VC単位の素片を利用して音声合成を行なう音声合成方法であって、音声辞書のサイズの低減させることができるとともに接続面での歪みを減少させることができる規則音声合成方法を提供することを目的とする。

## 【0010】

【課題を解決するための手段】この発明による規則音声合成方法は、HMnetを用いてCV、VC単位の素片を、音素単位より細かい合成単位に分割し、全ての合成単位にまたがる音素内ネットワークを生成し、得られた音素内ネットワークに基づいて入力音素列に対応する素片を選択するようにしたことを特徴とする。

【0011】音素内ネットワークは、たとえば、予め用意した個々のCV、VC単位の素片の始端部分から得られる複数の第1状態と、全てのCV、VC単位の素片の中央部分から得られる1つの第2状態と、個々のCV、VC単位の素片の終端部分から得られる複数の第3状態とによって構成されるHMnetの初期状態から、逐次状態分割法によって状態を分割するステップ、および任意の状態数まで分割されたHMnetと、個々のCV、V

C単位の素片とに基づいて、HMnetの各状態の代表素片を決定するステップによって生成される。

【0012】HMnetの各状態の代表素片を決定するステップは、たとえば、個々のCV、VC単位の素片と、任意の状態数まで分割されHMnetから、各学習用素片を各状態に分割し、各状態の素片の候補を作成するステップ、およびHMnetの全ての状態間の接続点に、HMnetの同じ音素からなる終端と始端の全ての組み合わせを加えたものを総接続点とし、総接続点での接続歪みの総和が最小となるような素片の組み合わせを各状態の素片候補の中から選択して、代表素片とするステップからなる。

【0013】

【発明の実施の形態】以下、図面を参照して、この発明の実施の形態について説明する。

【0014】〔1〕発明の特徴についての説明

【0015】この発明では、CV、VCを接続単位とした音声合成方法であって、隠れマルコフ網(HMnet)を用いて、全ての接続単位に対してネットワークを形成することにより、音声辞書の圧縮を行なうものである。

【0016】HMnetとは、音声認識で用いられる手法であり、HMMの学習時に、異なる音素や異音間で状態の共有を行なうことによって、全体のモデルのパラメータを少なくし、統計的に安定なモデルを推定しようとするものである(「鷹見、嵯峨山:『逐次状態分割法(SSS)による隠れマルコフネットワークの自動生成』日本音響学会平成3年度周期発表会講演論文集、2-5-1-(1991-9)」参照)。この発明では、HMnetの1状態に合成の1接続単位(合成単位)を割り当てた。

【0017】図1は、HMnetを用いてCV、VC単位の素片を、音素単位より細かい合成単位に分割することにより得られた音素内ネットワークの一例を示している。

【0018】各楕円は各状態を示し、各状態に合成の1接続単位(音素単位より細かい単位)が割り当てられている。楕円内のa-s, z, nは、この状態を[a]で始まり、[s], [z], [n]で終わる素片が通ることを示している。

【0019】〔2〕音素内ネットワークを生成する方法についての説明

【0020】以下、HMnetを用いてCV、VC素片をより細かく分割し、音素内ネットワークを生成する方法について説明する。

【0021】図2は、HMnetの初期状態を示している。

【0022】左一列は、学習用に用意した全てのCV、VC素片のうち、[a]で始まる素片を抽出し、その始端部分から計算した音響パラメータの平均値および分散

値からなる状態を[a\*]、[u]で始まる素片から同様に計算された状態を[u\*]、というようにして生成された複数の第1状態を表している。中央の状態は、全てのCV、VC素片の中心部分から計算した音響パラメータの平均値および分散値からなる1つの第2状態を表している。右一列は、個々のCV、VC素片の終端部分からそれぞれ計算した音響パラメータの平均値および分散値からなる複数の第3状態を表している。

【0023】状態の分割には、逐次状態分割法を用いた。音声認識に用いられる逐次状態分割法では、コンテキスト方向への分割を、音素の種類あるいは音素環境に基づいて行なっているが、この実施の形態では、音素の種類に係わらず状態を増やすことにした。

【0024】この理由は、音声合成時の音響パラメータの歪みが音素環境だけでなく、音韻環境などからも影響を受けることを考慮し、それらの要因に対しても状態を割り当てるためである。

【0025】逐次状態分割法によって、たとえば、図1に示すように、任意の状態数まで状態が分割されると、各状態の代表素片を次のようにして決定する。

【0026】(1)各学習用素片(CV、VC素片)と、任意の状態数まで分割されHMnetから、Viterbialgorithmを用いて、各学習用素片を各状態に分割し、各状態の素片の候補を作成する。

【0027】つまり、HMnetにおける任意の学習用素片に対応する経路における各状態での時間長の比に基づいて、当該学習用素片が各状態に分割される。同じ状態から同じ状態への遷移を伴わない状態についての時間長を1とした場合には、同じ状態から同じ状態への遷移を伴う状態については、同じ状態から同じ状態への遷移回数をnとすると、この状態での時間長は(1+n)となる。

【0028】(2)HMnetの状態遷移を元に、全ての状態間の接続関係を求める。

【0029】(3)上記接続関係にHMnetの同じ音素からなる終端と始端の全ての組み合わせを加えたものを総接続点とする。

【0030】(4)総接続点での接続歪みの総和が最小となるような素片の組み合わせを各状態の素片候補の中から選択し、代表素片とする。

【0031】上記(4)の代表素片の選択には、シミュレーテッドアニーリング(SA)を用いた。また、接続歪みは、次の数式1により計算した。

【0032】

【数1】

$$D = W_{F0}D_{F0} + W_{P0}D_{P0} + W_{cep}D_{cep}$$

【0033】 $D_{F0}$ 、 $D_{P0}$  および  $D_{cep}$  は、それぞれ接続する2つの合成単位の接続点での基本周波数の差、パワーの差およびケプストラム距離を表している。また、 $w_{F0}$ 、 $w_{P0}$  および  $w_{cep}$  は、それぞれ  $D_{F0}$ 、 $D_{P0}$  およ

よびDcepに乗ずる重み係数を表している。

【0034】〔3〕音声合成方法の説明

【0035】以下、図1の音素内ネットワーク(HMnet)を用いて、音声合成を行なう場合について説明する。

【0036】ここでは、「あさ」を合成する場合について説明する。「あさ」は、「a-s」の素片と、「s-a」の素片とにより合成される。

【0037】(1)まず、図3に示すように、「a-s」が通る経路を抽出する。

【0038】(2)次に、図4に示すように、「s-a」が通る経路を抽出する。

【0039】(3)そして、図5に示すように、これらの経路を接続する。

【0040】この場合、状態Aを通る経路と、状態Bを通る経路とがある。そこで、どちらの経路が接続歪みが小さいかを調べ、接続歪みが小さい方の経路を選択する。選択された経路上の各状態の代表素片を適当な高さ、長さ、大きさに修正した後、接続することによ

て、合成音声を生成する。

【0041】

【発明の効果】この発明によれば、予め用意したCV、VC単位の素片を利用して音声合成を行なう音声合成方法において、音声辞書のサイズの低減させることができるとともに接続面での歪みを減少させることができるようになる。

【図面の簡単な説明】

【図1】HMnetを用いてCV、VC単位の素片を、音素単位より細かい合成単位に分割することにより得られた音素内ネットワークの一例を示す模式図である。

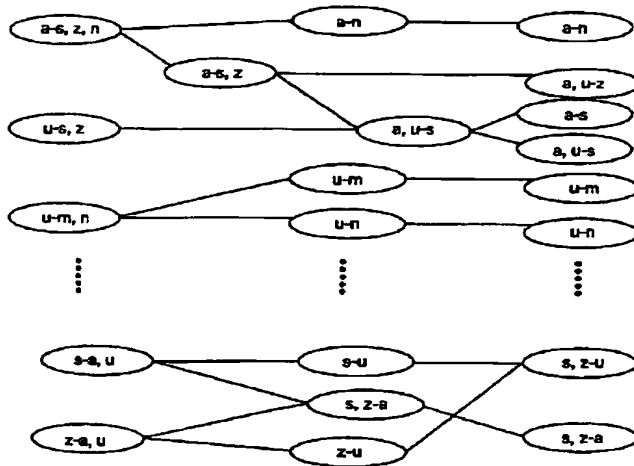
【図2】HMnetの初期状態を示す模式図である。

【図3】図1のHMnetのうち、「a-s」が通る経路を示す模式図である。

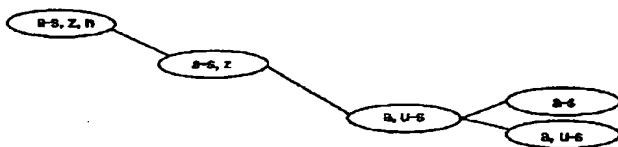
【図4】図1のHMnetのうち、「s-a」が通る経路を示す模式図である。

【図5】図3の経路と図4の経路とを接続することによって得られる経路を示す模式図である。

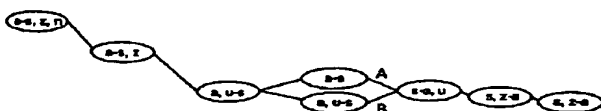
【図1】



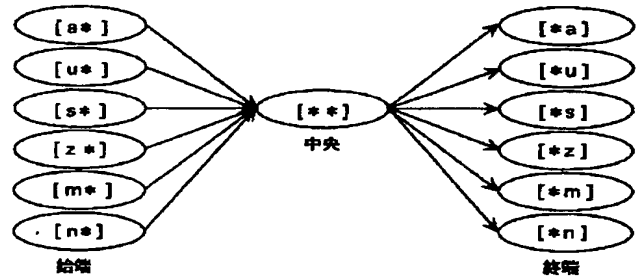
【図3】



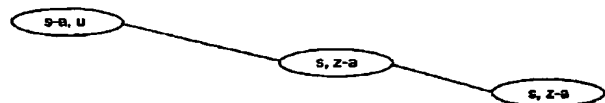
【図5】



【図2】



【図4】



フロントページの続き

(72)発明者 西田 秀治  
大阪府守口市京阪本通2丁目5番5号 三  
洋電機株式会社内

(72)発明者 大倉 計美  
大阪府守口市京阪本通2丁目5番5号 三  
洋電機株式会社内

(72)発明者 大西 宏樹  
大阪府守口市京阪本通2丁目5番5号 三  
洋電機株式会社内

Fターム(参考) 5D045 AA07